

Stochastic Frontier Production Function With Errors-In-Variables

Rajeev Dhawan
Department of Management
J. Mack Robinson College of Business
Georgia State University

&

Peter Jochumzen^{* †}
Department of Economics
Lund University

JEL Codes: C15;C21;D24

Keywords: Errors-In-Variables; Stochastic Frontier;
Technical Efficiency; Reliability Ratio

Abstract

This paper develops a procedure for estimating parameters of a cross-sectional stochastic frontier production function (SFPF) when input variables suffer from measurement errors. Specifically, we use Fuller's (1987) reliability ratio concept to develop an estimator for the model in Aigner, Lovell and Schmidt (1977). Our Monte-Carlo simulation exercise illustrates the direction and the severity of bias in the traditional MLE estimates of SFPF parameters. In contrast, the reliability ratio based estimator consistently estimates these parameters even under extreme degree of measurement errors. Additionally, traditional MLE estimates of firm level technical efficiency are severely biased rendering inter-firm efficiency comparisons infeasible. The seriousness of measurement errors in a practical setting is demonstrated by using data for publicly traded U.S. corporations.

* Corresponding author: The mailing address is Department of Economics, Lund university, Box 7082, 22007 LUND, Sweden. E-mail: peter.jochumzen@nek.lu.se, Phone: +46462228666.

† The authors wish to thank George Battese, Alok Bohara, Kurt Brännäs, David Edgerton, Kishore Gawande and Almas Heshmati for their numerous comments and suggestions that substantially improved this paper. We would also like to thank seminar participants at Lund University, University of New Mexico, conference participants at the econometric conference held at Stockholm School of Economics in May 1997, and the 3rd Biennial Georgia Productivity Workshop held at University of Georgia in October 1998 for their helpful comments and suggestions. The second author gratefully acknowledges financial support from The Swedish Council for Research in the Humanities and Social Sciences (HSFR), grant # F0353/97. However, any remaining errors are our own.

1. Introduction

Economic data is by nature non-experimental which compels the user at times to use either imperfectly measured variables, bad proxies or even outright dirty data. Consequently, *errors-in-variables* (EIV) is one of the most serious data problem facing an applied econometrician. It is well known that, for example, even relatively small measurement errors create quite severe biases in a linear regression model. EIV in the linear regression model has been analyzed quite thoroughly (see Fuller (1987) for a survey). Not much is known about correction methods for EIV problem in a non-linear regression framework¹.

Since the *stochastic frontier production function* (SFPF) framework was developed by Aigner, Lovell and Schmidt (1977) (ALS henceforth) and Meeusen and van den Broeck (1977), evaluating the efficiency of individual firms and industries has become popular with the increasing availability of firm-level data and growing capacity of computers to process them². The most common approach to estimate SFPF is to specify a deterministic, parametric production function common to all economic units. The *stochastic frontier* is then defined as the deterministic production function plus a random, symmetric, firm-specific error term. This frontier represents the largest possible production for the individual firm. Associated with each firm is a second, non-negative error term, denoted the *technical inefficiency* term. Total

¹ See Carroll, Ruppert and Stefanski (1995), although mostly directed to epidemiology, regarding measurement errors in nonlinear models. Another important reference is Wansbeek and Meijer (2000). A number of recent articles have suggested various methods for estimating general nonlinear measurement error models (for example Hsiao and Wang (2000), Newey (2001) and Li (2002)).

² Both economists and policy makers have made use of this trend as the notion, “frontier”, is associated with the theory of optimization, in addition to identifying factors that can explain relative efficiencies of economic units. A notable example is Caves and Barton (1990), who use the ALS framework to investigate factors that affect efficiency in U.S. manufacturing structure. A detailed list of studies that use the SFPF approach for efficiency measurement related issues can be found in a survey article by Greene (1997).

production for each firm is defined as the frontier minus the technical inefficiency³. Measurement errors for factors of production will produce inconsistent estimates of the production function as well as the shape and position of the frontier and the resultant technical efficiency estimate⁴. For example, incorrect estimate of returns to scale parameter has implications for growth theory and international economics, and at micro level it can result in faulty firm or industry efficiency analysis. Hence, this paper devises a methodology to investigate the severity of the bias introduced by measurement errors in the estimation of SFPF model.

The paper is organized as follows. Section II, develops a maximum likelihood point estimator for a simple SFPF model with measurement error, given sufficient *a priori* information to identify the model. The analysis is then extended to situations with weaker prior information where we develop “reasonable bounds” for the parameters of interest. Section III presents a Monte-Carlo simulation that illustrates a) the severity of the bias ignoring measurement errors and b) the consistency and small sample properties of the estimator developed in section II. An empirical example, based on U.S. firm-level data, is given in section IV. Section V concludes with a summary of our findings and some directions for future research.

³ Coelli, Rao and Battese (1998) and Kumbhakar and Knox Lovell (2000) contain detailed surveys of different econometric techniques for estimating SFPF and technical efficiency.

⁴ For example, a major source of measurement error in measuring flow of services from capital is the lack of information on the vintage nature of capital stock. Additionally, the labor usage figures reported by firms can also be mismeasured due to the non-availability of information on the skill levels of the labor force being used.

2. SFPF in Cross-Section

2.1. Basic Setup

Conceptually, there are many ways of specifying a SFPF and there are many ways of specifying an EIV model⁵. This paper will consider the basic setup of the SFPF model as in ALS with the most common assumptions on the structure of measurement errors used in the economics literature. An extended SFPF model with possibly other specifications on the measurement errors may be analyzed using the same basic procedure developed in this section. However, presenting solutions for all possible combinations is simply not feasible. For some extensions of the SFPF (such as allowing the mean of the technical inefficiency distribution to be non-zero) the modifications to the procedure is quite straightforward, whereas other extensions (such as using a CES type production function) is more difficult.

Let x_i denote the actual (unobserved) $k \times 1$ vector of inputs in log terms for firm i . We also observe an equal number of variables z_i which are related to x_i through $z_i = x_i + u_i$ where u_i denotes the $k \times 1$ vector of measurement errors (also in logs). The most common assumption in the EIV literature is that $x_i \sim \text{IIDN}(\mu, \Sigma_x)$ and $u_i \sim \text{IIDN}(0, \Sigma_u)$ and that x_i is independent of u_i . This implies that $z_i \sim \text{IIDN}(\mu, \Sigma_z)$ with $\Sigma_z = \Sigma_x + \Sigma_u$. This is the error model approach⁶. This

⁵ Since ALS the SFPF specification has been extended in many directions (see Greene (1997) for a detailed survey of the SFPF literature).

⁶ The other approach to modeling errors is the Berkson model where $x_i = z_i + v_i$, and v_i are the measurement errors. Here, $E(v_i | z_i) = 0$ which implies that the expectation conditional on observed z_i is zero, as opposed to the error models where the expectation conditional on actual x_i is zero. In addition, in the error model, the observed value is correlated with the measurement error while the actual value is correlated with the measurement error in the Berkson model. Which approach to use depends on what one believes is independent of the measurement errors: the true values of inputs or the observed values. In economics, the error model approach is completely dominating due to the way we believe that data is generated. The Berkson model is more appropriate in medicine and biology where the measured variable is often fixed by design.

model is non-calibrated in the sense that z_i is not specified to be related to x_i in a systematic manner⁷.

The cross sectional SFPF as in ALS may be written as:

$$y_i = \alpha + x_i' \beta + \varepsilon_i - \xi_i \quad (1)$$

where y_i is output in logs for the i 'th firm, ε_i is an IID random error variable which represents the statistical noise to the production structure and β is a $k \times 1$ vector of unknown parameters. ALS assume that $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ so that the maximum output a firm could produce using x_i is $\exp(\alpha + x_i' \beta + \varepsilon_i)$. Technical inefficiency is introduced as a positive random variable ξ_i . The most common assumption made in the literature is that ξ_i follows a truncated normal with mean zero (the positive half), i.e. $\xi_i \sim \text{IIDN}^+(0, \sigma_\xi^2)$ ^{8,9}. We also assume that the true factors (x_i), the measurement errors (u_i), the stochastic frontier error term (ε_i) and the technical inefficiency (ξ_i) are all independent random variables for all i .

Thus, the model considered in this paper is:

$$y_i = \alpha + x_i' \beta + \varepsilon_i - \xi_i \quad (2a)$$

$$z_i = x_i + u_i \quad (2b)$$

$$x_i \sim \text{IIDN}(\mu, \Sigma_x), u_i \sim \text{IIDN}(0, \Sigma_u), \varepsilon_i \sim \text{IIDN}(0, \sigma_\varepsilon^2), \xi_i \sim \text{IIDN}^+(0, \sigma_\xi^2) \quad (2c)$$

where x_i , u_i , ε_i and ξ_i are independent for all i . Often, one uses transformations of the parameters σ_ε^2 and σ_ξ^2 defined as following:

$$\lambda = \frac{\sigma_\xi}{\sigma_\varepsilon}, \sigma^2 = \sigma_\varepsilon^2 + \sigma_\xi^2 \text{ which implies, } \sigma_\xi^2 = \frac{\lambda^2 \sigma^2}{1 + \lambda^2}, \sigma_\varepsilon^2 = \frac{\sigma^2}{1 + \lambda^2} \quad (3)$$

⁷ An error model would be calibrated if it was specified that $z_i = \gamma_0 + \gamma_1 x_i + u_i$. As we have no reason, a priori, to believe that there is a systematic bias in the measurements of inputs, we have selected the non-calibrated model.

⁸ Other common assumptions are exponential (Meeusen and van den Broeck (1977) and also in ALS), gamma (Beckers and Hammond (1987), Greene (1990, 2003b)) and truncated normal with non zero mean (Stevenson (1980)). See Greene (1997) for a detailed discussion regarding the merits and shortcomings of these different distributional assumptions.

⁹ The variance of ξ_i is equal to $(1-2/\pi)\sigma_\xi^2$.

2.2. Identification and Reliability Ratios

It is clear that the model (2) is unidentifiable unless additional information is available. This occurs because we only observe z_i while x_i and u_i are unobserved, implying that Σ_u and Σ_x cannot be identified separately. One possibility is to use instrumental variables¹⁰. The procedure developed in this paper will not require any instrumental variables.

In order to progress further we need some additional information. We will first assume that we have precise information on the *reliability ratios* of the factors of production. This assumption will be relaxed later. The reliability ratios are defined as follows:

$$\pi_j = \text{Var}(x_{ij})/\text{Var}(z_{ij}) \quad j = 1, \dots, k \quad (4a)$$

$$\pi_{jm} = \text{Cov}(x_{ij}, x_{im})/\text{Cov}(z_{ij}, z_{im}) = \text{Cov}(x_{ij}, x_{im})/[\text{Cov}(x_{ij}, x_{im}) + \text{Cov}(u_{ij}, u_{im})] \\ j = 1, \dots, k \text{ and } m = 1, \dots, k, j \neq m \quad (4b)$$

where π_j is the (traditional) reliability ratio associated with variable j . Note that $0 \leq \pi_j \leq 1$ and π_j is equal to one if there are no measurement errors for variable j . Typically, one specifies the covariance matrix of the measurement errors u to be diagonal (see for example Klepper and Leamer (1984)). If this is the case, the reliability ratios π_j 's are all we need for identifying the model since all π_{jm} will be equal to one. However, we would like to allow for covariances between the different measurement errors, u_{ij} and u_{im} , which then warrants the introduction of π_{jm} . π_{jm} is the ratio of the true (unobserved) covariance between two variables to the observed covariance and will be called the *covariance reliability ratio*.¹¹ Note that if one assumes that the

¹⁰ Bound *et al* (2001, Section 3.3) discusses nonlinear measurement error models using instrumental variables. In the linear case, the estimates will be consistent if the instrument is uncorrelated with the measurement error. This is *not* the case for nonlinear models, which means that it is significantly more difficult to use instrumental variables in nonlinear models. Another disadvantage is the problem of finding and justifying such instruments for factors of production. If an instrument actually does exist, it is an act of faith to assume that the instrument's reduced correlation with the error term (in this context, its lower measurement error compared to the independent variable's) will more than compensate for the decreased precision (due to the non perfect correlation between the instrument and the variable).

¹¹ This is an extension of Fullers reliability ratio which, as far as we know, has not been used before.

measurement errors of different variables are uncorrelated, then all the π_{jm} 's are equal to one^{12,13}.

Since x_i and u_i are independent, it follows that:

$$\text{Var}(u_{ij}) = (1 - \pi_j)\text{Var}(z_{ij}) \quad j = 1, \dots, k \quad (5a)$$

$$\text{Cov}(u_{ij}, u_{im}) = (1 - \pi_{jm})\text{Cov}(z_{ij}, z_{im}) \quad j = 1, \dots, k \text{ and } m = 1, \dots, k, j \neq m. \quad (5b)$$

It is possible to collect all the reliability ratios into one symmetric matrix Π where:

$$\Pi = \begin{bmatrix} \pi_1 & \cdots & \pi_{1k} \\ \vdots & \ddots & \vdots \\ \pi_{k1} & \cdots & \pi_k \end{bmatrix}$$

With this notation, we may write, $\Sigma_x = \Pi \cdot \Sigma_z$ and $\Sigma_u = (\mathbf{1} - \Pi) \cdot \Sigma_z$ where the notation “ \cdot ” means element by element multiplication (the Hadamard product) and where $\mathbf{1}$ is a $k \times k$ matrix of ones.

2.3. Maximum Likelihood Estimation when Π is known

In this section we derive consistent estimates of α , β , σ_ε^2 , σ_ξ^2 , or equivalently, α , β , σ^2 and λ , in model (2) conditional on the reliability ratios Π . With no measurement errors, there are two possible solutions, maximum likelihood or OLS. OLS applied to (2a) yields consistent estimates of $\alpha - E(\xi_i) = \alpha - (\pi/2)^{1/2}\sigma_\xi$ and β . A consistent estimate of σ_ξ may be derived from the OLS residuals (see Olson, Schmidt and Waldman (1980)) which in turn will give us a consistent estimate of α (corrected OLS or COLS). COLS is, however, less efficient than MLE. With measurement errors, OLS will no longer give us consistent estimates of β or of $\alpha - E(\xi_i)$. Thus, we will derive the likelihood function associated with the model in (2). The results in the rest of this section are all conditional on Π .

¹² Also, unless the measurement errors are negatively correlated, the π_{jm} 's will be less than one.

¹³ If the variables x_{ij} and x_{im} are uncorrelated, we define π_{jm} to be equal to one.

The joint distribution of the observations on y_i and z_i will involve all the parameters given by $\omega' = (\mu', \text{vech}(\Sigma_z)', \alpha, \beta', \lambda, \sigma^2)$. We can write the log-likelihood of the parameters as¹⁴:

$$\ell(\omega) = \sum_{i=1}^n \ln f(y_i, z_i | \omega) = \sum_{i=1}^n \ln f_1(z_i | \omega_1) + \sum_{i=1}^n \ln f_2(y_i | z_i, \omega_1, \omega_2) = \ell_1(\omega_1) + \ell_2(\omega_1, \omega_2) \quad (6)$$

where $\omega_1' = (\mu', \text{vech}(\Sigma_z)')$ and $\omega_2' = (\alpha, \beta', \lambda, \sigma^2)$. We use the method of limited information maximum likelihood (LIML) to maximize the likelihood function¹⁵. First, we use the sample moments of z_i to estimate μ and Σ_z (which is equivalent to maximizing $\ell_1(\omega_1)$). We then maximize $\ell_2(\omega_1, \omega_2)$ over ω_1 using these sample moments.

To calculate the second-step likelihood we need to find the conditional distribution of y given z . Let v be the random variable $x'\beta$ and again define $e_i = \varepsilon_i - \xi_i$ as the compound residual.

Then by considering the joint density of y and v we have:

$$f_{y|z}(y|z) = \int_v f_{y,v|z}(y, v|z) dv = \int_v f_{y|v,z}(y|v, z) f_{v|z}(v|z) dv \quad (7)$$

Because we condition on z and $x'\beta$, we have:

$$\begin{aligned} f_{y|v,z}(y|v, z) &= f_{e|v,z}(y - \alpha - v|v, z) \\ &= f_{e|v}(y - \alpha - v|v) = f_e(y - \alpha - v) \end{aligned} \quad (8)$$

The last equality follows because e is independent of z and $x'\beta$. By substituting (8) into (7), we get:

$$f_{y|z}(y|z) = \int_v f_{v|z}(v|z) f_e(y - \alpha - v) dv \quad (9)$$

where the integral is a single one over all possible values of $x'\beta$. Since x and z are normal, $v | z$ is also normal. Straightforward application of the results for a conditional density of a multivariate normal gives the expected value and variance of $x'\beta | z$ ¹⁶

¹⁴ Assume that x , and therefore z , is weakly exogenous with respect to $(\alpha, \beta, \lambda, \sigma^2)$.

¹⁵ Estimating these parameters using full information maximum likelihood (FIML) could not be done as the likelihood function was too flat, mostly due to the elements in Σ_z , to allow us to find the maximum.

¹⁶ Since $E(x_i'\beta) = \mu\beta$ and $\text{Cov}(x_i'\beta, z_i) = \Sigma_x\beta$ we have

$$\begin{pmatrix} x_i'\beta \\ z_i \end{pmatrix} \sim N \left[\begin{pmatrix} \mu'\beta \\ \mu \end{pmatrix}, \begin{pmatrix} \beta'\Sigma_x\beta & \beta'\Sigma_z \\ \Sigma_x\beta & \Sigma_z \end{pmatrix} \right]$$

$$E(v | z) = \mu' \beta + (z - \mu)' \Sigma_z^{-1} \Pi \cdot \Sigma_z \beta \quad (10a)$$

and

$$\text{Var}(v | z) = \beta' (\Pi \cdot \Sigma_z - \Pi \cdot \Pi \cdot \Sigma_z) \beta = \beta' (\Pi \cdot \Sigma_z \cdot \Pi^c) \beta \quad (10b)$$

where $\Pi^c = 1 - \Pi$.

Thus, the density $f_{v|z}$ can be written as:

$$f_{v|z}(v|z) = \frac{1}{\sqrt{2\pi\beta' (\Pi \cdot \Sigma_z \cdot \Pi^c) \beta}} \exp \left[-\frac{[x' \beta - \mu' \beta - (z - \mu)' \Sigma_z^{-1} \Pi \cdot \Sigma_z \beta]^2}{2\beta' (\Pi \cdot \Sigma_z \cdot \Pi^c) \beta} \right] \quad (11)$$

Since the density of e_i may be written as (see Weinstein (1964)):

$$f_e(e_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \text{erfc} \left[\frac{\lambda e_i}{\sqrt{2\sigma^2}} \right] \exp \left[-\frac{e_i^2}{2\sigma^2} \right] \quad (12)$$

where λ and σ^2 are defined in (3) and $\text{erfc}(x)$ is defined as $1 - \text{erf}(x)$ where the error function $\text{erf}(x)$ is defined as:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

The conditional density of y_i given z_i is given by:

$$f(y_i | z_i) = \int_v \frac{1}{2\pi\sqrt{\sigma^2\beta' (\Pi \cdot \Sigma_z \cdot \Pi^c) \beta}} \text{erfc} \left[\frac{\lambda(y_i - \alpha - v)}{\sqrt{2\sigma^2}} \right] \times \exp \left[-\frac{(y_i - \alpha - v)^2}{2\sigma^2} - \frac{[v - \mu\beta - (z_i - \mu)' \Sigma_z^{-1} \Pi \cdot \Sigma_z \beta]^2}{2\beta' (\Pi \cdot \Sigma_z \cdot \Pi^c) \beta} \right] dv \quad (13)$$

and the second step likelihood function ℓ_2 is given by

$$\ell_2(\hat{\omega}_1, \omega_2) = \sum_{i=1}^n \ln f_2(y_i | z_i, \hat{\omega}_1(z), \omega_2) =$$

Then using the formula for the conditional distribution of $x_i' \beta$ given z_i we get $x_i' \beta | z_i \sim N[\mu' \beta + (z_i - \mu)' \Sigma_z^{-1} \Sigma_x \beta, \beta' (\Sigma_x - \Sigma_x \Sigma_z^{-1} \Sigma_x) \beta]$. This using $\Sigma_x = \Pi \cdot \Sigma_z$ could then be written as $N[\mu' \beta + (z - \mu)' \Sigma_z^{-1} \Pi \cdot \Sigma_z \beta, \beta' [\Pi \cdot \Sigma_z \cdot (1 - \Pi)] \beta]$.

$$\sum_{i=1}^n \ln \int_v \frac{1}{2\pi\sqrt{\sigma^2 \beta'(\Pi \cdot \hat{\Sigma}_z \cdot \Pi^c) \beta}} \operatorname{erfc} \left[\frac{\lambda(y_i - \alpha - v)}{\sqrt{2\sigma^2}} \right] \times \exp \left[-\frac{(y_i - \alpha - v)^2}{2\sigma^2} - \frac{[v - \hat{\mu}\beta - (z_i - \hat{\mu})' \hat{\Sigma}_z^{-1} \Pi \cdot \hat{\Sigma}_z \beta]^2}{2\beta'(\Pi \cdot \hat{\Sigma}_z \cdot \Pi^c) \beta} \right] dv \quad (14)$$

It is possible to show that this integral will not simplify unless every reliability ratio in Π is equal to one in which case the density above will converge to the density of e . However, it is possible to evaluate the likelihood function numerically for any matrix Π ¹⁷.

Note that the estimated covariance matrix of $(\alpha, \beta', \lambda, \sigma^2)$ at the second stage will no longer be consistent. However, by using the results of Murphy and Topel (1985), the adjusted asymptotic covariance matrix of the second-step estimates ω_2 (V_2^*) has to be calculated as following¹⁸:

$$V_2^* = V_2 + V_2(CV_1C - RV_1C - CV_1R)V_2 \quad (15)$$

where V_2 is the unadjusted second-step covariance matrix, V_1 is the asymptotic covariance matrix of $\hat{\omega}'_1 = (\hat{\mu}', \operatorname{vech}(\hat{\Sigma}_z)') = (\bar{z}', \operatorname{vech}(S_z)')$. Also, C and R , following Murphy and Topel (1985) who establish consistency of LIML under the usual regularity conditions, are defined as following¹⁹:

$$C = E \left[\begin{pmatrix} \frac{\partial \ell_2}{\partial \omega_2} \\ \frac{\partial \ell_2}{\partial \omega'_1} \end{pmatrix} \right]$$

$$R = E \left[\begin{pmatrix} \frac{\partial \ell_2}{\partial \omega_2} \\ \frac{\partial \ell_1}{\partial \omega'_1} \end{pmatrix} \right]$$

¹⁷ We need to evaluate n integrals each time we calculate the value of the likelihood function which is not a problem for a fast computer. Note that the likelihood function will simplify further if all the reliability ratios are equal to one common value.

¹⁸ See Edgerton and Jochumzen (2003) for a discussion of how to calculate the covariance matrix. In the simulation studies as well as in the empirical example, we found that the correction to V_2 was very small and we present the unadjusted covariance matrix.

¹⁹ The gradient vectors are difficult to calculate analytically so we used numerical derivatives to calculate C and R .

2.4. *Bounds on Reliability Ratio*

The previous section raises an important issue: it would be very useful if economic data was accompanied by reliability ratios. If we knew not just the data but how *reliable* the data was, it would be possible to derive consistent point estimates of the SFPF. When reliability information is lacking, we tend to analyze the data *as if* the reliability ratio was 100%, and, as we will see in the simulation study, if the actual reliability is less, the point estimates will be severely inconsistent.

Reliability ratios may be known in certain cases. Bound *et al* (2001) has a review of over 100 studies concerning measurement errors. A third of the studies give information which can easily be transformed into reliability ratios, while others can be used to derive these ratios under further assumptions.²⁰ In addition, some studies even give information concerning calibration and correlations between the measurement errors and the true variables.

However, for most practical problems and in near future, we will have to contend with data with little or no reliability information. Maximizing (14) for a given Π will provide us with a mapping from Π to an estimate of the SFPF: $\Pi \rightarrow \hat{\omega}_2$. If we specify *reasonable bounds* on Π , we can derive reasonable bounds on $\hat{\omega}_2 = (\hat{\alpha}, \hat{\beta}', \hat{\lambda}, \hat{\sigma}^2)$. This procedure is similar to Klepper and Leamer (1984) which identifies consistent bounds for β in a multiple regression model. Although it would be possible to derive consistent bound in the SFPF setting, we have decided not to do so. There are basically three reasons for using reasonable instead of consistent bounds. First, the consistent bounds tend to be very wide, often unbounded, and some estimates inside the

²⁰ The information, which is usually in the form of correlations, is most often found in studies concerning earnings, hours worked, benefits and education.

bounds may be associated with very unrealistic measurement errors. Second, There is no mapping between a particular estimate inside the consistent bounds and the associated Π . Thus, there is no way of knowing whether a particular consistent estimate is reasonable. Third, although prior information can be incorporated into the consistent bounds analysis, this can only be done in a non-intuitive manner using lower bounds on the correlations between explanatory variables *and* upper bounds on the R^2 of the regression equation.

Fortunately the data and the model itself will impose bounds on the Π -matrix as some Π -matrices will give rise to negative estimates of the variance of ε_i .²¹ This can easily be illustrated in the case of one variable input ($k = 1$) production function. Combining equations (2a) and (2b) one gets:

$$y_i = \alpha + \beta z_i + \varepsilon_i - \beta u_i - \xi_i \quad (16)$$

where the error term is now composed of three parts. Since βu_i is normal, the skewness of the data will determine the relative share of the variance between $\varepsilon_i - \beta u_i$ on one hand and ξ_i on the other in equation (16). Given this division, the data will determine the size of the variance of $\varepsilon_i - \beta u_i$ (denoted by $\sigma_{\varepsilon-\beta u}^2 = \sigma_\varepsilon^2 + \beta^2 \sigma_u^2$), but not how $\sigma_{\varepsilon-\beta u}^2$ is shared between the two terms σ_ε^2 and $\beta^2 \sigma_u^2$. With no measurement errors, all the $\sigma_{\varepsilon-\beta u}^2$ variance can be attributed to the variance of the residual ε . As the reliability ratio π decreases (so that the variance of u increases), more of the $\sigma_{\varepsilon-\beta u}^2$ variance will be due to the variance in βu_i . In the extreme case *all* the variance of $\varepsilon_i - \beta u_i$ is due to measurement errors. In this case $\sigma_\varepsilon^2 = 0$ and $\sigma_{\varepsilon-\beta u}^2 = \beta^2 \sigma_u^2$. Attempting to decrease π further will result in negative estimate of σ_ε^2 . Thus, the lower limit for the reliability ratio is determined by the point where the estimate of $\sigma_\varepsilon^2 = 0$.

²¹ Note that we only need to worry about the variance of ε_i as any Π implies a positive semi-definite variance matrix of x , u and z . Also, as long as the data on output is skewed to the left, the estimated variance of ξ_i is positive.

In general, if we define $\hat{\omega}_2^\Pi$ as the estimated value of ω_2 given the reliability ratio matrix Π , the restriction that the estimate of $\sigma_\varepsilon^2 \geq 0$ is equivalent to $\hat{\omega}_2^\Pi \Sigma_u \hat{\omega}_2^\Pi = \hat{\omega}_2^{\prime\Pi} (1 - \Pi \times \Sigma_x) \hat{\omega}_2^\Pi \leq \sigma_{\varepsilon-u}^2$, which is the restriction for determining feasible values of Π . Additional bounds on the reliability ratios may be found if some other simplifying assumptions are made. Such assumptions will be discussed in section III when the simulation exercise is undertaken.

2.5. Technical Efficiency

The production function implicit in equation (2a) when written in level terms is:

$$Y_i = e^\alpha \left[\prod_{j=1}^k X_{ij}^{\beta_j} \right] e^{\varepsilon_i} e^{-\xi_i}$$

where the output Y_i and input X_{ij} are in levels or original units. Thus, the technical efficiency of firm i can be defined as $\varphi_i = \exp(-\xi_i)$ ²² and it may be interpreted as the percentage of maximum possible output achieved when the residual ε_i is zero. This involves the technical inefficiency term ξ_i , which is an unobservable random variable. With no measurement errors and if α and β are known, even then only the difference $e_i = \varepsilon_i - \xi_i$ is recoverable. Hence, the best predictor of firm-specific technical efficiency is the expected value of $(\varphi_i | e_i)$. Jondrow et al. (1982) compute $E(\xi_i | e_i)$ as the predicted technical *inefficiency*. This is based on the approximation that $\xi_i = -\ln \varphi_i \approx 1 - \varphi_i$. We prefer to avoid this approximation and calculate directly the conditional expected value of φ_i , as suggested by Battese and Coelli (1988), which is²³:

$$E[\exp(-\xi_i) | e_i] = \exp \left\{ \frac{(2e_i + \sigma_\varepsilon^2) \sigma_\xi^2}{2\sigma^2} \right\} \left\{ 1 - \Phi \left[\frac{(e_i + \sigma_\varepsilon^2) \lambda}{\sigma} \right] \right\} \left\{ 1 - \Phi \left[\frac{e_i \lambda}{\sigma} \right] \right\}^{-1} \quad (17)$$

²² Since $\xi_i \geq 0$, the technical efficiency will always be between zero and one. A firm is defined as fully efficient or located on the frontier if $\xi_i = 0$ in which case the technical efficiency measure is equal to one.

²³ Note that this expression converges to $\exp(\hat{e}_i)$ as $\sigma_\varepsilon^2 \rightarrow 0$ which is what we should expect since \hat{e}_i in this case estimates $-\xi_i$.

Replacing e_i by the residual $\hat{e}_i = y_i - \hat{\alpha} - x_i' \hat{\beta}$ will give us an estimator of the technical efficiency of firm i . The mean or average technical efficiency of firms is straightforward to derive as:

$$E[\exp(-\xi_i)] = E(\varphi) = 2\exp(\sigma_\xi^2/2)[1-\Phi(\sigma_\xi)] \quad (18)$$

When measurement errors are present, the expression for the expected value of the mean technical efficiency, $E[\exp(-\xi_i)]$, is the same as in equation (18) since the distribution of ξ_i is unaffected. However, the expression for the technical efficiency for firm i is no longer valid. With measurement errors, the compound residual will be given by $e_{i*} = \varepsilon_i - u_i'\beta - \xi_i$ (see equation (16)) instead of $e_i = \varepsilon_i - \xi_i$ as in the case of no measurement errors. Since $u_i'\beta$ is normal, the expression for $E[\exp(-\xi_i) | \hat{e}_i]$ in equation (17) will be valid if we replace ε_i by $\varepsilon_i - u_i'\beta$ and redefine σ^2 and λ as following:

$$\sigma_*^2 = \sigma_\varepsilon^2 + \beta'\Sigma_u\beta + \sigma_\xi^2 \text{ and } \lambda_* = \frac{\sigma_\xi}{\sqrt{\sigma_\varepsilon^2 + \beta'\Sigma_u\beta}} \quad (19)$$

Consequently, the expression for conditional technical efficiency under errors will then be:

$$E[\exp(-\xi_i) | \hat{e}_{i*}] = \exp\left\{\frac{(2\hat{e}_{i*} + \sigma_\varepsilon^2 + \beta'\Sigma_u\beta)\sigma_\xi^2}{2\sigma_*^2}\right\} \left\{1 - \Phi\left[\frac{(\hat{e}_{i*} + \sigma_\varepsilon^2 + \beta'\Sigma_u\beta)\lambda_*}{\sigma_*}\right]\right\} \left\{1 - \Phi\left[\frac{\hat{e}_{i*}\lambda_*}{\sigma_*}\right]\right\}^{-1} \quad (20)$$

3. Simulation Study

3.1. Setup

This section compares the estimator for the cross-sectional SFPF developed in the previous section (henceforth called the EIVML estimator) with the traditional ML estimator on simulated data. The aim is twofold. We would like to investigate the *necessity* of the EIVMLE estimator by analyzing the bias introduced by measurement errors in estimating the production function parameters and the resulting technical efficiency estimates using traditional techniques. Second, we investigate the *properties* of the EIVMLE estimator by checking how it

fares in small samples. The model that we choose to simulate is a Cobb-Douglas production function with two inputs, capital and labor. The choice of only two inputs was motivated by the desire to be as similar to our empirical example presented in section IV, where the data allows for identification of only two broadly defined category of inputs: total capital and total labor. In addition, the basic parameters for simulation are chosen so as to closely mimic the actual data analyzed in section IV. The starting point of the simulation is the following model specification:

$$\ln(Y_i) = \alpha + \beta_K \ln(K_i) + \beta_L \ln(L_i) + \varepsilon_i - \xi_i \quad (21a)$$

$$\text{where } \ln(\underline{K}_i) = \ln(K_i) + \ln(U_{K_i}) \text{ and } \ln(\underline{L}_i) = \ln(L_i) + \ln(U_{L_i}) \quad (21b, 21c)$$

where K_i and L_i are actual but unobserved amount of capital and labor of firm i and \underline{K}_i and \underline{L}_i are the measured counterparts. U , ε and ξ are defined as in section II and, σ_K^2 denote the variance of $\ln(\underline{K}_i)$, σ_L^2 the variance of $\ln(\underline{L}_i)$ and σ_{KL} the covariance between $\ln(\underline{K}_i)$ and $\ln(\underline{L}_i)$. We define π_K and π_L as the reliability ratios of capital and labor and π_{KL} as the covariance reliability ratio.

Now, a slight modification of the model in (21) by writing it in per-capita terms is preferred. This is achieved by subtracting $\ln(L_i)$ from both sides of (21a) and subtracting (21c) from (21b). There are three advantages of doing this. First, it is easier to find the maximum of the likelihood function when regressing $\ln(Y/L)$ on $\ln(K/L)$ and $\ln(L)$ instead of regressing $\ln(Y)$ on $\ln(K)$ and $\ln(L)$. Second, the parameter of $\ln(L)$ will directly estimate the degree of departure from the constant returns to scale. Third, the per-capita specification allows us to find bounds on the *feasible* reliability ratios relatively easily as we discuss later in the next sub-section. Thus, after the transformation the model in equations (21a, b, c) can be written as:

$$\ln(Y_i / L_i) = \alpha + \beta_K \ln(K_i / L_i) + (\beta_L + \beta_K - 1)\ln(L_i) + \varepsilon_i - \xi_i. \quad (22a)$$

$$\ln(\underline{K}_i / \underline{L}_i) = \ln(K_i / L_i) + \ln(U_{K_i} / U_{L_i}), \quad \ln(\underline{L}_i) = \ln(L_i) + \ln(U_{L_i}) \quad (22b, 22c)$$

or equivalently as:

$$y_i = \alpha + x_i \gamma + \varepsilon_i - \xi_i. \quad (23a)$$

$$z_i = x_i + u_i. \quad (23b)$$

where $y_i = \ln(Y_i / L_i)$, $x_i = [\ln(K_i / L_i), \ln(L_i)]$, $z_i = [\ln(\underline{K}_i / \underline{L}_i), \ln(\underline{L}_i)]$, $u_i = [\ln(U_{Ki} / U_{Li}), \ln(U_{Li})]$ and $\gamma = [\beta_K, (\beta_L + \beta_K - 1)]^{24}$. Here we define π_1 as the reliability ratio of $\ln(K/L)$, π_2 as the reliability ratio of $\ln(L)$, which is simply π_L , and π_{12} as the covariance reliability ratio of $\ln(K/L)$ and $\ln(L)$.

It is not possible to present results for all possible combinations of the reliability ratios. First, we restrict $\pi_{KL} = \pi_L$. This is motivated by the small correlation between z_1 and z_2 (-0.03 in our data set). If $\text{Cov}(z_1, z_2) = \text{Cov}(x_1, x_2) = 0$ we must have $\pi_{KL} = \pi_L^{25}$. Second, we restrict $\pi_K = \pi_L$. The reason is that, given π_L , setting $\pi_K = \pi_L$ will produce the most extreme estimates of the parameters. Setting $1 \leq \pi_K < \pi_L$ will produce estimates similar to those with $\pi_K = \pi_L$ but with a smaller π_L . We denote this common value of π_L , π_K and π_{KL} by $\hat{\pi}$. This will also imply that $\pi_1 = \pi_2 = \pi_{12} = \hat{\pi}^{26}$. Further, we set $\text{Var}(\varepsilon_i) = 0.2$ and $\text{Var}(\xi_i) = 0.8$ (which implies that $\sigma^2 = 1$ and $\lambda = 2$) and $\alpha = 1.7$, $\gamma_1 = 0.6$ and $\gamma_2 = 0.1$ implying increasing returns to scale feature of the production structure i.e. $\beta_K = 0.6$ and $\beta_L = 0.5$.

3.2. Results: Parameter Estimates

Each simulation round consisted of 500 observations to estimate the parameters and this was repeated 100 times. Table 1 presents the averages and the standard deviations of the estimated parameters for the traditional MLE and the EIVMLE for different reliability ratios. In the table estimates of σ_ε^2 and σ_ξ^2 are derived from the estimates of σ^2 and λ using expressions defined in equation (3). The most striking result of the simulation study is the severe downward

²⁴ Note that x_i and u_i are independent since $\ln(K_i / L_i)$ and $\ln(U_{Ki} / U_{Li})$ are independent.

²⁵ $\text{Cov}(z_1, z_2) = \text{Cov}[\ln(\underline{K}_i / \underline{L}_i), \ln(\underline{L}_i)] = \text{Cov}[\ln(\underline{K}_i), \ln(\underline{L}_i)] - \text{Var}[\ln(\underline{L}_i)] = \sigma_{KL} - \sigma_L^2 = 0 \Rightarrow \sigma_{KL} = \sigma_L^2$. $\text{Cov}(x_1, x_2) = \text{Cov}[\ln(K_i / L_i), \ln(L_i)] = \text{Cov}[\ln(K_i), \ln(L_i)] - \text{Var}[\ln(L_i)] = \pi_{KL} \cdot \sigma_{KL} - \pi_L \cdot \sigma_L^2 = 0 \Rightarrow \pi_{KL} = \pi_L$

²⁶ $\pi_1 = (\pi_K \sigma_K^2 + \pi_L \sigma_L^2 - 2\pi_{KL} \sigma_{KL}) / (\sigma_K^2 + \sigma_L^2 - 2\sigma_{KL})$, $\pi_{12} = (\pi_{KL} \sigma_{KL} - \pi_L \sigma_L^2) / (\sigma_{KL} - \sigma_L^2)$

bias in the traditional ML estimate of γ_1 and γ_2 as the common reliability ratio falls. This implies that (where $\gamma_1 = \beta_K$ and $\gamma_2 = \beta_L + \beta_K - 1$) we *underestimate the elasticity of capital while we overestimate that of labor* when there are measurement errors. For example, in the simulated data the elasticity of capital was 0.6 while that of the labor was 0.5. With 80% reliability in the data, the capital elasticity is underestimated by almost one-fifth, and for 70% reliability ratio the estimates are completely reversed: 0.4175 for capital and 0.6499 for labor! Thus, the biases are quite severe and clearly show the need for a procedure that consistently estimates the elasticity parameters under even reasonable degree of measurement errors.

Results reported in table 1 also imply that the traditional MLE tends to underestimate the return to scale parameter γ_2 . Therefore, if one wants to test for increasing returns, the traditional MLE does a poor job whereas the EIVMLE will pick out true increasing returns even for a 70% reliability ratio. Table 1 also shows that the traditional MLE based λ estimate is biased downward and σ^2 is biased upwards. The combined effect of these two on the estimate for the variance of technical efficiency (σ_ξ^2) is that it seems to be estimated consistently, whereas the variance of the residual (σ_ε^2) is biased upwards. This is not surprising as the measurement errors being normally distributed, will be captured in the σ_ε^2 term, thus biasing it upwards, leaving the σ_ξ^2 estimate to be unaffected. Thus, there will be an upward bias in estimated σ^2 and a downward bias in estimated λ using traditional ML. Hence it is clear in table 1 that both σ^2 and λ are well estimated by the EIVML.

3.3. Results: Technical Efficiency

In table 2, column 2 presents the mean technical efficiency estimates for different reliability ratios where the true expected value of the unconditional technical efficiency using equation (18) and $\sigma_\xi^2 = 0.8$ is 0.5536. Columns 3 and 4 contain estimates of the expected value

of the technical efficiency based on traditional ML and EIVML estimates from table 1. Here the traditional ML estimator does as well as the EIVML estimator when it comes to estimating the mean value of the technical efficiency. This happens because both estimators produce estimates of σ_{ξ}^2 that are identical to the third decimal, and this is the only parameter that determines the average technical efficiency (see equation (21)). Hence, if you are only interested in the mean technical efficiency, you may just as well use the traditional MLE, even if the data suffers from measurement errors, as long as these are normally distributed.

Next, we analyze technical efficiency of firm i once we have estimated the residual for that firm. In each simulation round we compare the true technical efficiencies to the estimated technical efficiencies calculated using the two estimators. This comparison was done by calculating the average absolute deviation between the true technical efficiency and the corresponding traditional ML and EIVML estimates. This then results in two numbers for each simulation round. The means and standard deviations of these 100 simulation rounds are presented in table 3. It is clear from the table that EIVMLE is much more successful at estimating technical efficiency of an individual firm. The average absolute deviation between the true value and the traditional MLE rises as the severity of the measurement errors increases, unlike for the EIVMLE where the absolute deviation stays about the same.

In the last column of table 3, we report another test that illustrates the benefits of adjusting for measurement errors when estimating firm-specific efficiency. The test is for what percentage of the 500 observations the EIVMLE technique results in an estimate closer to the true value in comparison to the traditional MLE. Even with 90% reliable data, you get better firm-specific efficiency estimates in about 95% of the cases by using EIVMLE.

To summarize, the traditional ML estimator is seriously biased when it comes to estimating the elasticity coefficients of labor and capital under measurement errors. It is also a poor choice if you want to estimate the technical efficiency for a particular firm. However, both estimators will estimate the mean technical efficiency level very well.

4. Empirical Example

4.1. *The Data*

In this section we examine the impact of measurement errors on SFPF estimates of a production structure in actual data. We draw a cross-section of firms from the COMPUSTAT industrial data files maintained by Standard and Poor. These files consist of all the publicly traded firms on the U.S. stock exchanges and provide information on balance sheet components, cash flow and income statements and other relevant financial information. The frequency of reporting is annual²⁷. We chose the year 1988 for our analysis as it provided the most number of firms with relevant information²⁸. The number of employees a firm employs is the labor use (L_i). Standard practice is to define labor use in terms of hours worked but this information is not available in COMPUSTAT. As we don't know the proportion of skilled versus unskilled workers as well as their quality level, this imparts a source of measurement error to our labor use variable. To calculate the output of a firm or the value added Y_i , the cost of goods was subtracted from the sales figure²⁹.

²⁷ This is also the database analyzed by Dhawan and Gerdes (1997), Dhawan (2001), and Dhawan and Prabhu (1999) for productivity measurement issues, and covers the time-period 1970-1989.

²⁸ We could have chosen the year 1989 which is the terminal year of the database. Because of non-reporting of relevant information by quite a number of firms, the highest number of firms with usable information was present in 1988. Another reason for choosing 1988 was the fact that this year was characterized by a stable economic environment, especially the inflation situation and financial market volatility.

²⁹ Because the reporting procedure for the cost of goods component contains labor expenses, a component of the value added by a firm, the labor expense component was added to the above calculation. Since not every firm reports this item as an expense separate from cost of goods, this correction dropped the number of firms that could ultimately be used in the analysis.

To complete the value added calculations, total inventories were added to the above measure. The measure of capital K_i is the book value of total assets of a firm³⁰. Thus, we have full information to estimate the production structure, and accompanying level of technical efficiency for 484 firms.

4.2. Production Structure

The model that we estimate is identical to the one considered in the simulation study (see (23)). Tables 4a, 4b and 4c provide the necessary summary statistics for the data variables. In particular note that the covariance between z_1 and z_2 is almost zero. As before, $\pi_K = \text{Var}(\ln(K)) / \text{Var}(\ln(\underline{K}))$ and $\pi_L = \text{Var}(\ln(L)) / \text{Var}(\ln(\underline{L}))$ are the reliability ratios of capital and labor respectively, while π_{KL} is the ‘‘covariance reliability’’ ratio equal to $\text{Cov}(\ln(K), \ln(L)) / \text{Cov}(\ln(\underline{K}), \ln(\underline{L}))$. As explained in the simulation section, it is reasonable to set $\pi_{KL} = \pi_L$ if $\text{cov}(z_1, z_2)$ is close to zero which is the case here. Also, we will only consider cases where $\pi_K = \pi_L = \hat{\pi}$ ³¹, which, as discussed in the simulation study, will make the reliability ratio of $\ln(K/L)$ as well as the covariance reliability ratio of $\ln(K/L)$ and $\ln(L)$ equal to $\hat{\pi}$. Based on the summary statistics in table 4c, we can derive consistent estimates of the expected value and the variance of z . Given a particular reliability ratio $\hat{\pi}$, we can then find consistent estimates of the expected value and variance of x as well as of the variance of u and these are:

$$\hat{\mu}_x = \begin{pmatrix} 4.93 \\ 1.15 \end{pmatrix} \quad \hat{\Sigma}_x = \hat{\pi} \begin{pmatrix} 1.31 & -0.03 \\ -0.03 & 6.90 \end{pmatrix} \quad \hat{\Sigma}_u = (1 - \hat{\pi}) \begin{pmatrix} 1.31 & -0.03 \\ -0.03 & 6.90 \end{pmatrix}$$

³⁰ Using total assets as a proxy for productive, physical capital requires qualifications. First, this measure of assets includes the current investment component of a firm. Second, this measure includes cash and other short term liquid investments which may not be appropriate measures of physical capital. A justification for using this measure is the theoretical models and empirical evidence that extend the notion of production structure by incorporating the effects of liquidity and borrowing constraints [for e.g. see Gertler and Hubbard (1988), Gertler (1988) and references contained within].

³¹ Again, setting π_K to values different from π_L did not reveal anything interesting.

Thus, given this information and the discussion regarding reliability ratio bounds in section IID, the lowest possible value for the reliability ratio is 0.86. Any value lower than that will produce a negative estimate for σ_ε^2 .

4.3. *Parameter Estimates*

Table 5 presents the estimates of the parameters in equations (23a) and (23b) using three techniques: OLS, the traditional MLE and the EIVMLE estimator developed in this paper. The first row presents the estimates when simple OLS technique is used which can be characterized as estimating an *average* production function. As is well known, with no measurement errors, OLS will provide us with consistent but inefficient estimates of γ ; an inconsistent estimate of α and no estimates for σ_ε^2 and σ_ξ^2 . With measurement errors even the OLS estimate for the parameter γ is inconsistent. In the second row the MLE based estimates are presented. Rows 3 to 9 display the estimates using the EIVMLE technique based on likelihood function from equation (6). Each row provides a set of estimates for a particular common reliability ratio. These results should be interpreted as following: *If* the reliability ratio of labor and capital is 0.94 (say for example), then the consistently estimated coefficients are in this row. Based on these estimates for α , β , σ^2 and λ , we then can derive estimates for the elasticity of capital and labor (β_K and β_L) as well as the variances of ε and ξ (σ_ε^2 and σ_ξ^2) presented in table 6.

A number of interesting but not surprising results, given our simulation experience, are apparent from Tables 5 and 6. First, MLE underestimates the elasticity of capital. According to MLE, the return to capital is 0.6261 while it is as much as 0.7280 using the EIVMLE technique and for the reliability ratio is 0.86. We also find that MLE estimates return to scale very well which then implies that it is over estimating the elasticity of labor. Second, as the reliability ratio decreases the estimated λ increases while estimated σ_ε^2 goes to zero. This happens because as

the reliability ratio decreases, the variance of $u\beta$ increases. This will happen at the expense of a decline in the variance of ε , and as it goes to zero, λ , which is equal to $\sigma_{\xi}^2/\sigma_{\varepsilon}$, will increase³². Third, we find that MLE estimates, σ_{ξ}^2 , the variance of ξ_i very well. This has important implications for the estimates of the technical efficiencies as discussed later in the next subsection. MLE also overestimates the variance in ε_i , which is natural, since it assumes no measurement errors. Finally, the estimate of the intercept using MLE is significantly higher than the OLS estimates. This comes as no surprise since OLS ignores the ξ term, and as the expected value of ξ is positive, it explains the difference.

4.4. Estimates of Technical Efficiencies

Given that we have 486 firms, it is not possible to present the estimates of technical efficiency for each firm. We begin first by considering the mean average technical efficiency under varying degree of measurement errors. This is presented in table 7. It is interesting to note that the average level of firm efficiency is almost independent of the assumption on measurement errors. The EIVML estimates are also close to the traditional ML estimate of the expected value of the unconditional technical efficiency. This happens because the only parameter that determines the distribution of the technical efficiencies, σ_{ξ}^2 , is almost identically estimated by both techniques regardless of the degree of measurement errors. At first, this may suggest that measurement errors are not an issue when it comes to technical efficiencies. However, as we know from the results of the simulation section, correction becomes very important when considering firm-specific efficiency estimates.

³² As a matter of fact, $\hat{\pi} = 0.86$ is a lower bound for the reliability ratios. There simply is not enough variation in the data to support more measurement errors than this. With $\hat{\pi} = 0.86$, the only disturbance to the model, except for the technical inefficiencies, are measurement errors as ε vanishes in this case.

To get an idea of the bias caused by measurement problems, we present technical efficiency estimates of the first ten firms in table 9. From this table, we note that the direction of the bias ignoring measurement errors can go both ways. This means that the estimated technical efficiency for one firm is biased upwards while it is biased downward for another. Correcting for measurement errors then becomes very important if one wants to compare technical efficiencies for different firms.

To explore this more, and to get an idea about how severe the problem could be, we ranked all the firms in the sample by their traditional ML based technical efficiency estimates. Then, as the reliability ratio was decreased, it was found that the relative ranking of the firms changed. For the reliability ratio 0.98 the maximum rank change was 23 on the upper side and 19 on the lower side. In addition, 50 percent of changes in ranks were between plus 2 to minus 2. For the lowest feasible reliability ratio of 0.86, 50% of the rank changes were within plus 15 and minus 16. For this particular reliability ratio the maximum rank change was 132 on the upper side and 131 on the lower side! In percentage terms the maximal change in firm level efficiency was 22% on the up side and 14% on the down side.

This is an important outcome since the technical efficiency estimate tells us what percentage of “frontier output” the firm is delivering. This precludes the researcher, who is not correcting for measurement errors, from establishing a comparative efficiency ranking of the firms in the sample as evident from the EIV estimate.

5. Summary and Conclusions

This paper investigates the impact measurement errors in inputs have on estimates of production function parameters and firm-specific technical efficiency estimates in a cross-

sectional SFPP setting. We first develop the methodology for estimating the standard cross sectional SFPP with measurement errors by using Fuller's reliability ratio concept. Next, our numerical simulation results show that the estimates (elasticity parameters) of the deterministic frontier, the distribution of the stochastic part of the frontier and the distribution of the technical inefficiency are very sensitive to the degree of measurement error. Our simulation results indicate that traditional MLE will bias the elasticity coefficient estimates and the returns to scale feature. These biases are quite severe and clearly demonstrate the need for a method that consistently estimates the production function parameters for even small degree of measurement errors.

The simulation exercise also shows that while traditional MLE overestimates the variance of the composite error term, it underestimates the skewness parameter with the result that the variance of the technical efficiency parameter is well estimated. Now, although the mean level of technical efficiency or average sample efficiency is unaffected by the presence of measurement errors, the firm-specific estimate of technical efficiency will be seriously biased as it depends upon the estimated skewness parameter.

Next, the practical applicability of the reliability ratio estimator developed in this paper is demonstrated by applying it to actual firm level data from the U.S. industrial sector. For this data set issues regarding returns to scale feature, elasticity coefficients and firm-specific technical efficiency are explored in detail. Here we demonstrate how the relative ranking of the firms by their technical efficiency estimates changes when the degree of measurement errors is increased. In addition, the percent change in the firm-specific technical efficiency levels from its ML estimate is quite severe when the degree of measurement error increases. This exercise has an implication for economic researchers who are engaged in inter-firm or inter-industry

comparisons – ignoring measurement errors will most likely lead to erroneous efficiency comparisons.

The analysis in this paper has been undertaken for cross-sectional SFPP model with Cobb-Douglas production structure and with fairly strong assumptions on the error structure that in many respects is very simplistic. Consequently, practical issues such as analyzing technical change over time and evolution of a firm's efficiency levels that requires a more general production structure (say Translog) in a panel setting are a subject matter of future research.

References

- Aigner, D.J., C.A.K. Lovell, and P. Schmidt. (1977). "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics* 6, 21-37.
- Battese, G.E. and T.J. Coelli. (1988). "Prediction of Firm-Level Technical Efficiencies With a Generalized Frontier Production Function and Panel Data," *Journal of Econometrics* 38, 387-399.
- Beckers, D. and C. Hammond. (1987). "A Tractable Likelihood Function for the Normal-Gamma Stochastic Frontier Model," *Economics Letters* 24, 33-38.
- Bound, J., Brown, C., and Mathiowetz, N. (2001), "Measurement Error in Survey Data," Chapter 59 in *Handbook of Econometrics, Volume 5*, Heckman, J. J., and Leamer, E. (eds), Elsevier, Amsterdam.
- Carroll, R.J., Ruppert D. and L.A. Stefanski. (1995). *Measurement Errors in Nonlinear Models*, Monographs on Statistics and Applied Probability 63, Chapman & Hall/CRC.
- Caves, R.E. and D.R. Barton. (1990). *Efficiency in U.S. manufacturing Industries*, Cambridge, Massachusetts: MIT Press.
- Coelli T.J., D.S.P. Rao and G. Battese. (1998). *An Introduction to Efficiency and Productivity Analysis*, Boston: Kluwer Academic Publishers.
- Dhawan, R. (2001). "Firm Size and productivity Differential: Theory and Evidence From a Panel of U.S. Firms," *Journal of Economic Behavior and Organization* 44, 269-293.
- Dhawan, R. and G. Gerdes. (1997). "Estimating Technological Change using a Stochastic Frontier Production Function Framework: Evidence From U.S. Firm-Level Data," *Journal of Productivity Analysis* 8, 431-446.

- Dhawan, R. and J. Prabhu. (1999). "The Competitive Advantage of Being Small: A Strategic Efficiency Versus Risk Approach," Cambridge University mimeograph.
- Edgerton, D. and P. Jochumzen. (2003). "Estimation in Binary Choice Models with Measurement Errors" Lund University, Department of Economics, WP 2003:4, July 2003.
- Fuller, W.A. (1987) *Measurement Error Models*, New York: Wiley.
- Gertler, M.L. (1988). "Financial Structure and Aggregate Economic Activity: An Overview," *Journal of Money, Credit and Banking* 20, 559-588.
- Gertler, M.L. and R.G. Hubbard. (1988). "Financial Factors and Business Fluctuations," in *Financial Market Volatility*, Federal Reserve Bank of Kansas City.
- Greene, W. (1990). "A Gamma Distributed Stochastic Frontier Model." *Journal of Econometrics* 46, 141-164.
- Greene, W. H. (1997). "Frontier Production Functions," in M. Hashem Pesaran and P. Schmidt (eds), *Handbook of Applied Econometrics, Vol II: Microeconomics*, Massachusetts: Blackwell Publishers.
- Greene, W. H. (2003a). *Econometric Analysis. Fifth Edition*, Prentice Hall, Upper Saddle River, New Jersey
- Greene, W. H. (2003b). "Simulated Likelihood Estimation of the Normal-Gamma Stochastic Frontier Function," *Journal of Productivity Analysis* 19, 179-190.
- Hsiao, C., and Wang, Q. K. (2000). "Estimation of Structural Nonlinear Errors-in-Variables Models by Simulated Least-Squares Method," *International Economic Review*, 41, 523-42.
- Jondrow, J., C.A.K. Lovell, I.S. Materov, and P. Schmidt. (1982). "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model," *Journal of Econometrics* 19, 233-238.

- Klepper, S., and E. E. Leamer. (1984). "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica* 52, 163-83.
- Kumbhakar, S.C and C.A.K. Lovell. (2000). *Stochastic Frontier Analysis*, Cambridge: Cambridge University Press.
- Li, T. (2002). "Robust and Consistent Estimation of Nonlinear Errors-in-Variables Models," *Journal of Econometrics*, **110**, 1-26.
- Meeusen, W.J. and J. van den Broeck. (1977). "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," *International Economic Review* 18, 435-444.
- Murphy K. and R. Topel. (1985). "Estimation and Inference in Two Step Econometric Models," *Journal of Business and Economic Statistics* 3, 370-379.
- Newey, W. K. (2001). "Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models," *Review of Economics and Statistics*, **83**, 616-27.
- Olson, J.A., P. Schmidt, and D.M. Waldman. (1980). "A Monte Carlo Study of Estimators of Stochastic Frontier Production Functions," *Journal of Econometrics* 13, 67-82.
- Pal, M., Neogi, C. and B. Ghosh. (1998). "Estimation of Frontier Production Function with Errors-In-Variables: An Illustration from Indian Industry," in Sr. Chakravarty, D. Coondoo and R. Mukherjee (eds.) *Quantitative Economics: Theory and Practice*, New Delhi: Allied Publishers Limited.
- Stevenson, R. (1980). "Likelihood Functions for Generalized Stochastic Frontier Estimation," *Journal of Econometrics* 13, 58-66.
- Wansbeek, T., and Meijer, E. (2000). *Measurement Error and Latent variables in Econometrics*, Elsevier, Amsterdam

Weinstein, M. (1964). "The Sum of Values from a Normal and a Truncated Normal Distribution," *Technometrics* 6, 104-105.

Table 1. Traditional Maximum Likelihood Estimates of SFPF for Simulated Data

	α	γ_1	γ_2	σ^2	λ	σ_{ξ}^2	σ_{ε}^2
True Value*	1.700	0.6000	0.1000	1.000	2.000	0.8000	0.2000
Traditional Maximum Likelihood Estimates							
$\hat{\pi}=1.00$	1.711 (0.07)	0.5998 (0.02)	0.1030 (0.02)	1.013 (0.13)	2.132 (0.39)	0.8242 (0.15)	0.1883 (0.04)
$\hat{\pi}=0.90$	1.691 (0.08)	0.5376 (0.02)	0.09289 (0.02)	1.055 (0.14)	1.757 (0.36)	0.7903 (0.18)	0.2644 (0.05)
$\hat{\pi}=0.80$	1.685 (0.09)	0.4778 (0.02)	0.08143 (0.03)	1.100 (0.14)	1.606 (0.35)	0.7848 (0.19)	0.3154 (0.06)
$\hat{\pi}=0.70$	1.676 (0.12)	0.4175 (0.02)	0.06741 (0.03)	1.121 (0.16)	1.494 (0.39)	0.7641 (0.22)	0.3568 (0.07)
EIV Method Maximum Likelihood Estimates							
$\hat{\pi}=1.00$	1.711 (0.07)	0.5998 (0.02)	0.1030 (0.02)	1.013 (0.13)	2.132 (0.39)	0.8242 (0.15)	0.1883 (0.04)
$\hat{\pi}=0.90$	1.691 (0.08)	0.5973 (0.03)	0.1032 (0.02)	0.9890 (0.14)	2.058 (0.49)	0.7903 (0.18)	0.1987 (0.05)
$\hat{\pi}=0.80$	1.685 (0.09)	0.5972 (0.03)	0.1018 (0.03)	0.9833 (0.14)	2.095 (0.62)	0.7848 (0.19)	0.1985 (0.06)
$\hat{\pi}=0.70$	1.676 (0.12)	0.5964 (0.03)	0.0963 (0.04)	0.9681 (0.17)	2.116 (0.85)	0.7641 (0.22)	0.2040 (0.08)

* The data was simulated from the model $y_i = \alpha + x_i\gamma + \varepsilon_i - \xi_i$ with $z_i = x_i + u_i$, $x_i \sim N(0, 2\hat{\pi})$, $u_i \sim N(0, 2(1-\hat{\pi}))$ where $\hat{\pi}$ is the common reliability ratio of log of labor, log of capital (and thus of log capital by labor). $\hat{\pi}$ is varied in the table and $\varepsilon \sim N(0, 0.2)$ and $\xi \sim N^+(0, 0.8)$. The standard errors are reported in parentheses.

Table 2. Mean Technical Efficiency And Reliability Ratio

Reliability ratio	Actual TE Value	Estimated TE (MLE Estimate)	Estimated TE (EIV Estimate)
1.00	0.5536	0.5496	0.5496
0.90	0.5536	0.5553	0.5553
0.80	0.5536	0.5562	0.5562
0.70	0.5536	0.5598	0.5598

Table 3. Comparing Firm-Specific Technical Efficiency Estimates

Reliability ratio	Average absolute deviation between EIVMLE and true value	Average absolute deviation between MLE and true value	Percentage won by EIVMLE
1.00*	N/A	N/A	N/A
0.90	0.0542 (1.8×10^{-3})	0.0755 (1.9×10^{-3})	94.47% (1.0%)
0.80	0.0463 (1.6×10^{-3})	0.125 (3.1×10^{-3})	99.27% (0.4%)
0.70	0.0406 (1.4×10^{-3})	0.161 (2.5×10^{-2})	99.72% (0.2%)

* For a reliability ratio of 1, MLE and EIVMLE will produce exactly the same estimates and the formulas for expected value of the conditional technical efficiencies will coincide. N/A implies not applicable here. The standard errors are reported in parentheses.

Table 4a. Transformed and Non-Transformed Data Variable Means

	$\ln(Y)$	$\ln(K)$	$\ln(L)$	$Y = \ln(Y/L)$	$Z_1 = \ln(K/L)$	$Z_2 = \ln(L)$
Mean:	5.55	6.08	1.15	4.40	4.93	1.15

Table 4b. Untransformed Data Variance and Covariance Matrix

	$\ln(Y)$	$\ln(K)$	$\ln(L)$
$\ln(Y)$	8.23	7.17	6.86
$\ln(K)$	7.17	8.14	6.86
$\ln(L)$	6.86	6.86	6.90

Table 4c. Transformed Data Variance and Covariance Matrix

	$Y = \ln(Y/L)$	$Z_1 = \ln(K/L)$	$Z_2 = \ln(L)$
Y	0.78	0.76	0.26
z_1	0.76	1.31	-0.03
z_2	0.26	-0.03	6.90

Table 5. SFPPF Parameter Estimates: OLS, traditional MLE And EIVMLE*

	α	γ_1	γ_2	σ^2	λ
OLS	1.5036 (0.33)	0.5781 (0.07)	0.04132 (0.03)	N/A	N/A
MLE	1.9195 (0.10)	0.6261 (0.02)	0.00071 (0.01)	0.7146 (0.07)	2.7072 (0.37)
EIVMLE $\hat{\pi}=0.98$	1.8565 (0.10)	0.6389 (0.02)	0.00073 (0.01)	0.7041 (0.07)	2.8904 (0.46)
EIVMLE $\hat{\pi}=0.96$	1.7910 (0.10)	0.6522 (0.02)	0.00074 (0.01)	0.6931 (0.07)	3.1270 (0.54)
EIVMLE $\hat{\pi}=0.94$	1.7226 (0.11)	0.6661 (0.02)	0.00076 (0.01)	0.6817 (0.06)	3.4486 (0.68)
EIVMLE $\hat{\pi}=0.92$	1.6512 (0.11)	0.6806 (0.02)	0.00078 (0.01)	0.6698 (0.06)	3.9188 (0.99)
EIVMLE $\hat{\pi}=0.90$	1.5767 (0.11)	0.6957 (0.02)	0.00079 (0.01)	0.6573 (0.06)	4.6979 (1.66)
EIVMLE $\hat{\pi}=0.88$	1.499 (0.12)	0.7115 (0.03)	0.00081 (0.01)	0.6443 (0.07)	6.376 (4.19)
EIVMLE $\hat{\pi}=0.86$	1.417 (0.14)	0.7280 (0.03)	0.00083 (0.01)	0.6307 (0.07)	18.51 (1.69)

* The standard errors are in parentheses and N/A means not applicable.

Table 6. Basic Production Structure Estimates

	β_K	β_L	σ_ξ^2	σ_ε^2
OLS	0.5781	0.4351	N/A	N/A
MLE	0.6261	0.3746	0.6288	0.0858
EIV $\hat{\pi}=0.98$	0.6389	0.3618	0.6288	0.0753
EIV $\hat{\pi}=0.96$	0.6522	0.3485	0.6288	0.0643
EIV $\hat{\pi}=0.94$	0.6661	0.3347	0.6288	0.0529
EIV $\hat{\pi}=0.92$	0.6806	0.3203	0.6289	0.0409
EIV $\hat{\pi}=0.90$	0.6957	0.3051	0.6288	0.0285
EIV $\hat{\pi}=0.88$	0.7115	0.2893	0.6288	0.0155
EIV $\hat{\pi}=0.86$	0.7280	0.2728	0.6289	0.00184

Table 7. Average Technical Efficiency For the Sample

	MLE ($\hat{\pi}=1$)	$\hat{\pi}=0.98$	$\hat{\pi}=0.96$	$\hat{\pi}=0.94$	$\hat{\pi}=0.92$	$\hat{\pi}=0.90$	$\hat{\pi}=0.88$	$\hat{\pi}=0.86$
Mean	0.6008	0.6008	0.6008	0.6008	0.6008	0.6012	0.6068	0.6180

Table 8. Predicted Firm Efficiency of the First 10 Firms

	MLE($\hat{\pi}=1$)	$\hat{\pi}=0.98$	$\hat{\pi}=0.96$	$\hat{\pi}=0.94$	$\hat{\pi}=0.92$	$\hat{\pi}=0.90$	$\hat{\pi}=0.88$	$\hat{\pi}=0.86$
Firm 1	0.753	0.759	0.765	0.771	0.777	0.783	0.788	0.793
Firm 2	0.487	0.487	0.487	0.487	0.488	0.489	0.489	0.492
Firm 3	0.885	0.886	0.888	0.890	0.891	0.891	0.891	0.891
Firm 4	0.933	0.933	0.932	0.931	0.930	0.928	0.926	0.923
Firm 5	0.810	0.814	0.818	0.822	0.826	0.829	0.832	0.834
Firm 6	0.827	0.829	0.832	0.834	0.836	0.838	0.839	0.840
Firm 7	0.281	0.275	0.271	0.266	0.262	0.258	0.250	0.250
Firm 8	0.615	0.623	0.630	0.638	0.647	0.655	0.664	0.673
Firm 9	0.714	0.712	0.710	0.707	0.704	0.701	0.697	0.694
Firm 10	0.596	0.595	0.594	0.593	0.592	0.592	0.591	0.591